DOCUMENT RESUME

ED 206 666 TM 810 561

AUTHOR Rubin, Donald L.
TITLE Psychometric and Pedagogical Validity in Large Scale Assessments of Oral Communication Skills.
PUB DATE [79]
NOTE 29p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Basic Skills; *Educational Assessment; Elementary Secondary Education; *Speech Communication; *Speech Tests; Test Construction; *Testing Programs; *Test Validity

ABSTRACT

Due to the recent Federal redefinition of "basic skills," large scale assessments of speaking and listening proficiencies will proliferate. In planning assessment procedures, educators must be sensitive not only to psychometric validity, but also to pedagogical validity: the effects of testing on curricular content and instructional strategies. Validity judgments are particular to procedures and user values. But ramifications of assessment characteristics of both types of validity can be identified. These characteristics include: (1) direct measures; (2) pure measures; (3) culturally pluralistic measures; and (4) contextually diverse measures. An illustrative measurement procedure is described. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED206666

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PSYCHOMETRIC AND PEDAGOGICAL VALIDITY IN LARGE SCALE ASSESSMENTS OF ORAL COMMUNICATION SKILLS

Donald L. Rubin

The University of Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D L. Rubin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

running head: large scale assessment

author's address; Departments of Language Education
and Speech Communication
125 Aderhold Hall
The University of Georgia
Athens, GA 30602

TM 810 561

Large Scale Assessment

Abstract

Due to the recent Federal Redefinition of "basic skills," large scale assessments of speaking and listening proficiencies will proliferate. In planning assessment procedures, educators must be sensitive not only to psychometric validity, but also to pedagogical validity - the effects of testing on curricular content and instructional strategies. Validity judgements are particular to procedures and user values. But ramifications of assessment characteristics both types of validity can be identified. These characteristics include (1) direct measures, (2) pure measures, (3) culturally pluralitic measures, and (4) contextually diverse measures. An illustrative measurement procedure is described.

The next several years are certain to witness increased demand for large scale assessments of oral communication skills. Language arts education will be influenced by the inclusion of speaking and listening in recent Federally mandated definitions of "basic skills."[1] Emphasis on basic skills, as popularly construed, is an offspring of the same social-educational movement that spawned competency based education. In the vision of this movement an important function of public schools is to certify minimum levels of student achievement. In this climate schools and teachers are held accountable for educational outcomes primarily in terms of test scores. At present there are few well established instruments for measuring speech communication competencies in evaluative, nonresearch contexts.[2] The manner in which the educational community develops and selects large scale procedures for assessing speaking and listening may have profound consequences on the future status of oral communication in the schools.

Several factors are salient with respect to test construction and adoption, among them monetary cost, time, amount of error contaminating measurements, and the utility or consequences of resulting certification decisions. This paper addresses itself to one such factor, considerations affecting the validity of oral communication assessments.

From a psychometric perspective, validity pertains to what can be legitimately inferred on the basis of a given assessment procedure. In general, three interdependent types of psychometric validity are recognized.[3] Content validity is determined by inquiring whether test items (including elicitation procedures and scoring criteria) represent the domain of knowledge or skill of interest. Criterion referenced validity is demonstrated empirically if an instrument discriminates among individuals in a manner

similar to results obtained from other related measurements (concurrent validity) or if the instrument forecasts test-takers' attributes, especially behavioral patterns, at some time after test administration (predictive validity). Inferences derived from an assessment procedure have construct validity if the test's rationale and operation are consistent with the network of conceptual relations of which the target knowledge or skill is in theory a part. Conducted by means of on-going empirical and logical analysis, construct validation directly queries, is this a test of what it purports?

The relevance of a testing procedure can be examined from a pedagogical, as well as psychometric point of view. In the present educational melieu test results may or may not have impact as feedback-utility for diagnoses and remediation of instructional weaknesses. But testing most surely has feedforward impact; classroom practices are altered to conform to test specifications. Simple power relations in educational societies demand that teachers "teach to the test." For better or for worse, instruction expands or contracts in accordance with available technologies of educational measurement.

It is unlikely for example that the vigorous support presently enjoyed by reading programs would be possible in the absence of advanced reading assessment methodologies. The adoption of sentence combining training in many quarters is motivated largely by the fact that its outcome, syntactic complexity, is easily assessed. This situation obtains despite research which disconfirms any direct relationship between syntactic complexity and judged quality of written composition.[4] Yet another instructive illustration of the effects of testing on instruction is the history of the Educational Testing Service's College Board English Composition Test. ETS

deleted original student writing samples from its test of writing ability when test constructors developed a multiple choice examination which was reliable, cost efficient, and highly predictive of writing sample ratings. However English teachers protested that the multiple choice testing was causing a reduction in student composition and increased classroom emphasis on items of usage, punctuation, and the like included in the test. Ultimately this pressure resulted in the reinstitution of the writing sample.[5] It is warrantable to conclude that the amount and type of public school instruction in speech communication is currently, and will continue to be, dependent upon the availability and character of large scale assessment procedures.

Pedagogical validity, then, pertains to what students would acquire as a result of instruction designed to optimize performance on a given assessment procedure. Two interdependent types of pedagogical validity may be distinguished. Curricular validity refers to the judged appropriateness of course content that is consistent with test demands. For example, if reconstructing a particular communication model is deemed tangential to effective speech, then a procedure which assesses ability to reconstruct that model has low curricular validity for instruction intended to promote effective speech. Instructional practices validity is a function of the judged appropriateness of teaching strategies used to implement course objectives. If it were determined, for example, that group problem solving is a functional teaching strategy for promoting effective speech, then an assessment which requires experience in group problem solving would have high instructional practices validity in this situation.

Test validation, whether in terms of pedagogical or psychometric

factors, is particularistic. Validity judgements hold only for specified populations, purposes, and procedures from which inferences are drawn.[6] Moreover, conceptualizations of communicative competence are divergent.[7] It is therefore difficult to draw conclusions about factors contributing to validity in oral assessments which are applicable over a range of testing needs and situations. These conclusions are best drawn by testing agencies' individual constituencies. Nevertheless, certain broad issues bearing upon determinations of psychometric and pedagogical validity in oral proficiency testing can be identified.

Direct Measures

Oral communication qua skill is a behavioral construct. Knowing about speech communication is not equivalent to being able to execute communication acts, though intellectual apprehension of pertinent variables presumably enhances skill. Mead attempted to construct a paper-and-pencil instrument which asked testees to choose appropriate communication strategies or to characterize likely outcomes in supplied situations. The product was of limited value as a standardized test.[8] Strategy choice is so highly context-dependent that it is unlikely that a supplied, hypothetical situation could provide rich enough information for choice among alternatives. Thus, Howie-Day found developmental trends in rationales used to justify choices among alternative strategies, but not developmental patterning in the choices, themselves.[9]

In a similarly indirect approach, cognitive and attitudinal pre-requisites to communication can be indexed, but measures of underlying traits may account for only small portions of the variance in communication performance. A case in point is the use of perspective-taking measures such as the Feffer Role Taking Task which McCaleb and Korman advocate as an

assessment tool.[10] However neither correlational nor intervention studies confirmed a strong dependency between this type of social cognitive measure and referential communication accuracy.[11]

It is altogether possible, however, that an indirect assessment of oral communication skills of adequate psychometric validity could be constructed. Still, matters of pedagogical validity interpose. Adopting such an instrument might undermine experiential learning and the acquisition of performance skills. Educators who value such practices and outcomes might be justified in rejecting indirect measures on this basis alone.

Indeed, the Speech Communication Association "Criteria for Evaluating Instruments and Procedures That Assess Speaking and Listening Skills" do demand evaluations which require students to demonstrate communication behaviors.[12] Among the disadvantages of performance measures of speech communication are cost and measurement error, although previous experience in writing evaluation suggests that these difficulties can be managed.[13] Prominent attacks on the validity of such performance measures arise because of inherent difficulties in controlling extraneous sources of variance in performances.

If speech communication skills are conceived as including invention of content material, then biases in favor of certain subject matter will be troublesome. For example, if students are asked to speak about "My Summer Vacation," then the speaker who has worked at an archeological dig in Greece will be at an advantage over one who just "hung around and played a little ball." This advantage is a source of psychometric invalidity, intruding variance irrelevant to any performance skill.

One approach to eliminating content differences as sources of variation is to constrain definitions of oral competencies to include only delivery

skills and thus excluding invention, organization, and stylistic choice. Under such constrained definitions it is proper to employ oral reading or memorized renditions of controlled texts as evaluation tools. The consequence of this testing procedure is to jeopardize pedagogical validity by encouraging reversion to narrowly elocutionary instruction. A second approach to at least minimizing effects of content differences is to simply instruct raters to attend more to effectiveness of presentation than to content. Only experimental study can reveal if such instruction is an effective technique. A third means of reducing variation due to response topic entails constructing communication tasks in which all or most necessary information is supplied. Referential communication accuracy exercises in which speakers encode supplied stimulus materials are of this type. A persuasive task that builds in necessary information is presented in a later section of this paper. A final technique for minimizing subject matter bias is to employ test items which can be presumed to be within the ken of testees. A question like, "Tell me about your favorite teacher here at school" ought to provide a more equivalent opportunity for discriminating on the basis of communication skills than an item which instead emphasizes differences in background knowledge and experience like, "What do you consider to be the most important event in professional athletics this past year?"

Assessments which sample naturalistic interaction would be accorded high psychometric validity. In naturalistic observation of performance, however, one again likely sacrifices the control and consistency needed for reliable judgements. Therefore many direct measures of communication will involve contrived or hypothetical contexts. Some research instruments ask subjects to report about their performance as in the instructions,

"What would you say to this person in order to get him to return your ball."[14] Regardless of their value as techniques of inquiry, such reports are not to be confused with techniques of performance evaluation. In contrast, other procedures ask subjects to role-play or perform as if they were in the specified interaction situation. Role-playing tasks can reflect naturalistic performance skills to the extent that they (1) supply sufficiently rich context including audience, purpose, and setting and (2) portray this context in a compelling enough manner to provide at least a pretext for ignoring the evaluative motivation for the exercise. A role-play task in which an examiner states, "Pretend I'm your friend and you want me to go to the movies with you," violates both these conditions. However a simulated employment interview for a job at a fast-food restaurant might provide a useful speech sample. An advantage of role-playing assessments is that for those who consider role-playing to be a useful instructional practice, such procedures have high pedagogical validity. Gene Sanford identifies a number of factors that can interfere with instructional role-playing and which may also have a bearing on role-playing for evaluative purposes.[15]

"Pure" Measures

Effective oral communication requires the intersection of verbal, social, and logical abilities. It is at the same time a motor and perceptual skill and is also influenced by attitudes. To search for a pure measure of communication competence is akin to the alchemists' attempts to isolate elemental fire. Nevertheless, and without expounding yet another definition of speech communication, speaking and listening are uniquely characterized by the spontaneous confluence of these subskills in the processes of conveying and extracting meanings. Measuring these processes constitutes an

especially elusive enterprise and consequently proposed assessment procedures may capture one or another of the more accessible components of communication skill, employing the channel of speech only incidentally.

One critique of currently available standardized listening tests is that they are little more than reading tests presented orally.[15] They fail to encompass the communicative nature of the listening process, stressing literal comprehension and ignoring such components of listening as utilizing paralinguistic cues, judging speaker attitudes, and forecasting content. Published tests of listening may in fact be measures of general verbal ability.

Similar criticisms may be leveled at any set of evaluation criteria that credit particular linguistic or stylistic features in an absolute fashion. Style is context dependent and the mark of communicative competence may indeed be flexibility, and not consistency, in the use of particular linguistic resources.[17] An illustration of this point is use of syntactic complexity as an index of oral proficiency as proposed by both Loban and McCaleb.[18] Though syntactic complexity is easily measured and is a useful indicator for many purposes, recent findings argue against simplistic age norm interpretations.[19] Jenson similarly found that speech context exerted greater impact on spoken syntactic complexity than did ability level.[20]

Use of Standard English pronunciations, lexicons, and grammars as a measure of oral communication competence is a related, albeit more controversial, subject. The rating instrument in use in the Gary, Indiana school district is an instance of an evaluation technique stressing Standard English dialects in an absolute fashion.[21] However not all occasions require Standard English. In some contexts, speakers may be penalized for

speaking it. Intelligibility is a rhetorical criterion of effectiveness. "Correctness" and gentility are prescriptive vehicles used primarily for social stratification. The two types of criteria are not necessarily equivalent. On the other hand, competent speakers do employ a formal style when a social situation is construed as of great interpersonal distance. Formality may be signalled by a number of features of which standard dialect is one, but which also include distinct enunciation, erect posture, and use of deferential terms. An assessment of oral communication skills need not be a test of spoken Standard English any more than it need be a test of general verbal ability or of syntactic complexity.

As evaluation procedures emphasizing particular sets of linguistic features raise issues of psychometric validity, so do they affect pedagogical validity. A testing emphasis on Standard English dialects, for example, could result in a classroom concentration on "remediating" nonstandard dialects. Emphasis on syntactic complexity could result in arhetorical sentence combining drill. Depending on a community's instructional objectives, these may be entirely defensible goals. But if educators wish to promote in students varied repertoires of communicative resources and wish to provide practice in contextually appropriate selection from those repertories, then qualitative standards for language cannot be absolute.[22]

Personality factors also merit discussion as variables potentially affecting the purity of oral communication measures. Traits like generalized communication apprehension and a disposition to tentativeness may result in negatively evaluated speech styles.[23] Rigid personalities are likely to perform poorly on certain listening tasks.[24] If the effects of these personality traits are conceptualized as extraneous contaminants of scores, then they can be independently measured and statistically partialled

out of proficiency indicators. The opposing point of view, that such personality factors are inherent to communication skill, can lead to far-reaching ramifications concerning the schools' responsibility for personality modification therapy. Yet a third position might hold that, while it is beyond the scope of public education to engage in deliberate mass therapy, it is proper to offer classroom instruction which imparts to students strategies for coping with dysfunctional personality traits in given situations.[25]

Culturally Pluralistic Measures

In oral language arts, as in no other subject area, educators are forced to confront cultural biases of the public schools. For it is through spoken language, more than any other behavior, that individuals project cultural identity and concommittant social status. For at least the past two decades educational policy has been subject to Sisyphean torture by the opposing forces of pluralistic egalitarianism and demands for quality control wherein the qualities defined as minimally acceptable tend to be middle class/technocratic. Schools have always operated to engineer the social order and it is not distorting to view the contemporary back to basics movement in this light as well.[26]

The equation of Standard English dialect use with communication competence demonstrates a particular cultural orientation, since some ethnic, regional, and socio-economic groups do not natively speak standard dialects (although all speakers display some degree of code-switching in accordance with parameters of communicative context). Other obvious sources of cultural bias in oral assessment include questions which call upon a range of experience limited to members of particular subcultures. Mead, by way of illustration, determined that minority group youngsters identified

the statement, "In order to get a good job you have to finish school," as a statement of fact more often than did Anglo students.[27]

Less detectable, but of greater potential impact, are sources of cultural bias due to differing cultural norms for what is to count as an appropriate response or what is construed at all as a communicative context. In certain blue-collar cultures only circumscribed types of self-disclosures are deemed appropriate, and those only under highly delineated circumstances.[28] Members of other cultures, on the other hand, may be promiscuous self-disclosers. When questioned by an authority figure Latino youngsters may avert their eyes as a sign of respect, while Anglo children may engage in eye-contact to demonstrate sincerety. Research suggests at least some social class differences in norms of verbal restraint as opposed to spontaneous elaboration of responses.[29]

Indeed, middle class children are trained in their home environments to perform in examination situations. Parents typically ask quasi-questions, questions which are clearly not requests for enlightenment since answers are patently known by the inquisitors. Or children are given practice in producing extended discourse primarily for the gratification of their elders. In sum, these children learn that speaking in the role of examinee is normal and rewarding.

This portrayal contrasts sharply with Philips' ethnography of communication in a Native American Indian community.[30] In this minority culture youngsters typically learn domestic skills by silently observing their elders, and practicing in private. Demonstrating a skill for purposes of formative evaluation is not normal. Rather, a child will perform the new task as a matter of course in its functional context after he or she is satisfied of mastery. Moreover, from early childhood

children are immersed in a peer culture with little basis for the role of examinee. Leadership is a shared and fluid role and it is unusual for a young member of this community to hold forth in speech for an extended period. For these reasons it seems inappropriate to engage in classroom recitation. Likewise, this culture provides no foundation for responding in a situation deliberately contrived as an examination of oral competence.

It is difficult to avoid the conclusion that no one oral assessment procedure can treat members of diverse cultures equivalently. Recognizing and choosing to accept selected cultural inequities inherent in communication testing may be a defensible posture. Recalling that validity judgements pertain to the uses of tests rather than to the tests themselves, interpreters are compelled to draw differing inferences from scores obtained from differing populations, however. Rendering interpretations, selecting assessment materials, criteria, and certification cut-off points are most sensibly conducted at as local a level as feasible and in accordance with community values.

Contextually Diverse Measures

The terms "basic skills" and "minimal competencies" connote abilities directly applicable to economic, political, biological, and perhaps familial survival. They do not translate readily to concern for self-concept, fulfillment, creativity, and psychic well-being. Consequently many competencey based oral communication objectives refer to "life roles" like giving and comprehending travel directions, interviewing for employment, and understanding television commercials.[31] Although not dichotomously opposed to life role objectives, speech communication instruction

has traditionally also encompassed humanistic and humanizing goals associated with the broad liberal arts of inquiry, self-knowledge, and autotelic participation in community. Only a modest minority of individuals has frequent occasion to deliver prepared orations once embarked upon their "life roles." It is not unlikely that a narrowly utilitarian testing domain could decimate attention to public oratory in crowded curricula. By the same token, assessment procedures which elicit extended nonspontaneous speech or which rely on criteria of voice, diction, and formal language style could as easily lead to excision of instruction in dyadic and small group interaction skills. To the extent that reduction in the scope of oral communication curricula is undesirable, tests which sample from a limited range of communication domains incur poor pedagogical validity.

Employing a battery of tasks representing a variety of communication contexts is a means of promoting a rich and varied speech communication curriculum. Increasing the number of test items in this manner also has salubrious effects on reliability of measurement. But this approach is also very costly. A more efficient system of equivalent pedagogical validity would be testee self-selection, or else random assignment to one of a variety of communication tasks. Rigorous field testing will be required in this case to ensure that the several procedures are, in fact, "parallel forms" and yield equivalent certification decisions. The question of equivalency is one of pedagogical as well as psychometric concern. For if students train to mastery on each of the tasks in classroom situations, the probability of measurement error due to test form is reduced.

An Illustrative Measure

The procedure described in this section is presently undergoing field

testing in conjunction with a competency based education demonstration project in the State of Georgia. It is presented here for illustrative purposes only, and not as a validated instrument proposed for adoption. This test is intended for tenth-grade administration as an equivalent form option to a simulated job interview. The task is also deliberately tied to pre-instruction in the communication context and the evaluation criteria. Students engage in some guided practice in similar communication tasks. To reduce expectancy effects and otherwise increase reliability, the test is administered and scored by a pool of trained teachers from the district who are assigned to test sites other than their instructional assignments.

This task is a role-play of a public hearing before a county commission. Figure 1 reproduces test materials given to students. It was selected to sample extended persuasive discourse in a public setting con-

Insert Figure 1 about here

sidered to be of life role significance. The task is highly structured, providing a degree of content consistency. It specifies purpose, audience, and choice of topic. While testees are apprised of some information of evidential value, they must decide upon a position, invent arguments, and are free to go beyond the information given.

In administering the procedure, the room is arranged with a podium inset into the audience area and facing a row of seats occupied by student mock-commissioners. The "Agenda" is read aloud and students may also refer to their printed copies. Testees speak in random order,[32] and the audience is encouraged to applaud after each presentation. Video-tape apparatus records the performances from an angle.

Evaluation criteria appear in Figure 2. The rationale for this scoring

Notes

[1]"Part 162- Basic Skills Improvement," Federal Register, 44 (April 27, 1979), 25151. For a compilation of large scale assessment programs in progress see K.L. Brown, P. Backlund, J. Gurry, and F. Jandt, State of the Art and Recommendations for Instrument Development, Vol. I, (Boston: Massachusetts Board of Education Bureau of Research and Assessment, 1979), 50-67.

[2]Brown, Backlund, Gurry, and Jandt, p. 39. A useful list of mostly research instruments appears in C. Larson, P. Backlund, M. Redmond, and A. Barbour, Assessing Functional Communication (Falls Church, Virginia: Speech Communication Association, 1978).

[3]Standards for Educational and Psychological Tests (Washington: American Psychological Association, 1974).

[4]H. Rosen, "An Investigation of the Effects of Differentiated Writing Assignments on the Performance in English Composition of A Selected Group of 15/16-Year-Old Pupils," Diss., University of Lordon, 1969.

[5]The English Composition Test With Essay (Princeton: Educational Testing Service, 1979), 1.

[6]L.J. Cronbach, "Test Validation," in Educational Measurement ed. R.L. Throndike (Washington: American Council of Education, 1971), 447.

[7]C.E. Larson, "Problems in Assessing Functional Communication," Communication Education, 27 (1978), 304-309.

[8]N.A. Mead, "Issues Related to a National Assessment of Speaking and Listening Skills," Paper presented at the Speech Communication Association Annual Meeting, Washington, D.C., December, 1977.

[9]A.M. Howie-Day, "Metapersuasion: The Development of Reasoning About Persuasive Strategies," Diss., University of Minnesota, 1977.

[10]J. McCaleb and D.L. Korman, "A Measure of Communication Development," The English Journal, 67:7 (1978), 41-45.

[11]G.L. Piché, M.L. Michlin, F.L. Johnson, and D.L. Rubin, "Relationships Between Fourth-Graders' Performances on Selected Role-Taking Tasks and Referential Communication Accuracy Tasks," Child Development, 46 (1975), 965-969; M.J. Chandler, S. Greenspan, and C. Barenboim, "The Assessment and Training of Role-Taking and Referential Communication Skills in Institutionalized Emotionally Disturbed Children." Developmental Psychology, 10 (1974), 546-554.

[12]Spectra, 15:5 (1979), 5.

[13]S. Hudson and L.R. Veal, Assessing Writing Competence Through Writing Samples, Studies in Language Education, Report No. 35 (Athens, Ga.: University of Georgia Department of Language Education, 1979).

[14]J.G. Delia and R.A. Clark, "Cognitive Complexity, Social Perception, and the Development of Listener-Adapted Communication in Six-, Eight-, Ten-, and Twelve-Year-Old Boys," Communication Monographs, 44 (1977), 326-345.

[15]G. Sanford, "Why Role Playing Fails," The English Journal, 63:9 (1974), 50-54.

[16]C.M. Kelly, "An Investigation of the Construct Validity of Two Commercially Published Listening Tests," Speech Monographs, 32 (1965), 139-143.

[17]D. Hymes, "Competence and Performance in Linguistic Theory," in Language Acquisition: Models and Methods eds. R. Huxley and E. Ingram (New York: Academic Press, 1979), 3-28.

[18]W. Loban, Language Development: Kindgergarten Through Grade Twelve, Research Report No. 18 (Urbana, Ill.: National Council of Teachers of English, 1976), i; J. McCaleb, "Measuring Oral Communication," English Education, 11 (1979), 41-47.

[19]M. Crowhurst, "On the Misinterpretation of Syntactic Complexity Data," English Education, 11 (1979), 91-97.

[20]J.M. Jenson, "A Comparative Investigation of the Casual and Careful Oral Language Styles of Average and Superior Fifth Grade Boys and Girls," Research in the Teaching of English, 7 (1973), 338-350.

[21]Oral Proficiency Program: Secondary Schools, 1977-1978, (Gary, In.: Gary Community School Corporation, 1977), 76.

[22]These goals are expressed in R.R. Allen and B.S. Wood, "Beyond Reading and Writing to Communication Competence," Communication Education, 27 (1978), 286-292.

[23]see, for example, E.A. Lind and W.M. O'Barr, "The Social Significance of Speech in the Courtroom," in Language and Social Psychology eds. H. Giles and R.N. St. Clair (Oxford, England: Basil Blackwell, 1979), 66-87.

[24]S.W. Lundsteen, Listening: Its Impact on Reading and the Other Language Arts (Urbana, Ill.: National Council of Teachers of English, 1979), 63-75.

[25]W.J. Fremouw and M.D. Scott, "Cognitive Restructuring: An Alternative Method for the Treatment of Communication Apprehension," Communication Education, 28 (1979), 129-133.

[26]G.L. Piché, "Class and Culture in the Development of the High School English Curriculum, 1880-1900," Research in the Teaching of English, 11 (1977), 17-27.

[27]Mead, p. 20.

[28]G. Philipsen, "Speaking 'Like a Man' in Teamsterville: Cultural Patterns of Role Enactment in an Urban Neighborhood," Quarterly Journal of Speech, 61 (1975), 13-22.

[29]F. Williams and R.C. Naremore, "On the Functional Analysis of Social Class Differences in Modes of Speech," Speech Monographs, 36 (1969), 77-102.

[30]S.V. Philips, "Acquisition of Rules for Appropriate Speech Usage," Monograph Series on Language and Linguistics, No. 23 (Washington: Georgetown University Press, 1970), 77-96.

[31]For example, R.E. Bassett, N. Whittington, and A. Staton-Spicer, "The Basics in Speaking and Listening for High School Graduates: What Should Be Assessed?" Communication Education, 27 (1978), 293-303. For an inciteful attack on communication life skills approaches to English education see J.J. Rouse, "Knowledge, Power, and the Teaching of English," College English, 40 (1979), 473-491.

[32]Speaking order effects may pose a serious defect in this procedure. Data are currently being analysed to determine if this is so.

[33]R. Lloyd-Jones, "Primary Trait Scoring," in Evaluating Writing eds. C.R. Cooper and L. Odell (Urbana, Ill.: National Council of Teachers of English, 1977), 33-66.

Insert Figure 2 about here

system is similar to the Primary Trait method used in the National Assessment of Educational Progress' writing assessment.[33] That is, criteria are specifically keyed to the rhetorical dimensions of the communication task rather than to universally applicable communication competencies. Since the scores are intended as diagnostic feedback to students as well as for certification decisions, and in order to enhance inter-rater reliability, an attempt is made to concretely describe indicators for each ordered level of performance. Two judges independently rate the videotaped presentations according each criterion equal weight.

Many language arts educators believe that large scale assessment and competency certification is antithetical to effective instruction. Certainly communication does not thrive in distinctively evaluative climates. Moreover, speech communication skills do not readily lend themselves to the types of reductionistic formulations that ease the labor of test construction. However by carefully considering broad issues of both psychometric and pedagogical validity in concert, it may be possible to create large scale assessment procedures whose ultimate product is a heightening of the quality of speech communication in our society.

Notes

[1]"Part 162- Basic Skills Improvement," Federal Register, 44 (April 27, 1979), 25151. For a compilation of large scale assessment programs in progress see K.L. Brown, P. Backlund, J. Gurry, and F. Jandt, State of the Art and Recommendations for Instrument Development, Vol. I, (Boston: Massachusetts Board of Education Bureau of Research and Assessment, 1979), 50-67.

(10)

[2]Brown, Backlund, Gurry, and Jandt, p. 39. A useful list of mostly research instruments appears in C. Larson, P. Backlund, M. Redmond, and A. Barbour, Assessing Functional Communication (Falls Church, Virginia: Speech Communication Association, 1978).

[3]Standards for Educational and Psychological Tests (Washington: American Psychological Association, 1974).

[4]H. Rosen, "An Investigation of the Effects of Differentiated Writing Assignments on the Performance in English Composition of A Selected Group of 15/16-Year-Old Pupils," Diss., University of London, 1969.

[5]The English Composition Test With Essay (Princeton: Educational Testing Service, 1979), 1.

[6]L.J. Cronbach, "Test Validation," in Educational Measurement ed. R.L. Throndike (Washington: American Council of Education, 1971), 447.

[7]C.E. Larson, "Problems in Assessing Functional Communication," Communication Education, 27 (1978), 304-309.

[8]N.A. Mead, "Issues Related to a National Assessment of Speaking and Listening Skills," Paper presented at the Speech Communication Association Annual Meeting, Washington, D.C., December, 1977.

[9]A.M. Howie-Day, "Metapersuasion: The Development of Reasoning About Persuasive Strategies," Diss., University of Minnesota, 1977.

[10]J. McCaleb and D.L. Korman, "A Measure of Communication Development," The English Journal, 67:7 (1978), 41-45.

[11]G.L. Piché, M.L. Michlin, F.L. Johnson, and D.L. Rubin, "Relationships Between Fourth-Graders' Performances on Selected Role-Taking Tasks and Referential Communication Accuracy Tasks," Child Development, 46 (1975), 965-969; M.J. Chandler, S. Greenspan, and C. Barenboim, "The Assessment and Training of Role-Taking and Referential Communication Skills in Institutionalized Emotionally Disturbed Children." Developmental Psychology, 10 (1974), 546-554.

[12]Spectra, 15:5 (1979), 5.

[13]S. Hudson and L.R. Veal, Assessing Writing Competence Through Writing Samples, Studies in Language Education, Report No. 35 (Athens, Ga.: University of Georgia Department of Language Education, 1979).

[14]J.G. Delia and R.A. Clark, "Cognitive Complexity, Social Perception, and the Development of Listener-Adapted Communication in Six-, Eight-, Ten-, and Twelve-Year-Old Boys," Communication Monographs, 44 (1977), 326-345.

[15]G. Sanford, "Why Role Playing Fails," The English Journal, 63:9 (1974), 50-54.

[16]C.M. Kelly, "An Investigation of the Construct Validity of Two Commercially Published Listening Tests," Speech Monographs, 32 (1965), 139-143.

[17]D. Hymes, "Competence and Performance in Linguistic Theory," in Language Acquisition: Models and Methods eds. R. Huxley and E. Ingram (New York: Academic Press, 1979), 3-28.

[18]W. Loban, Language Development: Kindgergarten Through Grade Twelve, Research Report No. 18 (Urbana, Ill.: National Council of Teachers of English, 1976), 1; J. McCaleb, "Measuring Oral Communication," English Education, 11 (1979), 41-47.

[19]M. Crowhurst, "On the Misinterpretation of Syntactic Complexity Data," English Education, 11 (1979), 91-97.

[20]J.M. Jenson, "A Comparative Investigation of the Casual and Careful Oral Language Styles of Average and Superior Fifth Grade Boys and Girls," Research in the Teaching of English, 7 (1973), 338-350.

[21]Oral Proficiency Program: Secondary Schools, 1977-1978, (Gary, In.: Gary Community School Corporation, 1977), 76.

[22]These goals are expressed in R.R. Allen and B.S. Wood, "Beyond Reading and Writing to Communication Competence," Communication Education, 27 (1978), 286-292.

[23]see, for example, E.A. Lind and W.M. O'Barr, "The Social Significance of Speech in the Courtroom," in Language and Social Psychology eds. H. Giles and R.N. St. Clair (Oxford, England: Basil Blackwell, 1979), 66-87.

[24]S.W. Lundsteen, Listening: Its Impact on Reading and the Other Language Arts (Urbana, Ill.: National Council of Teachers of English, 1979), 63-75.

[25]W.J. Fremouw and M.D. Scott, "Cognitive Restructuring: An Alternative Method for the Treatment of Communication Apprehension," Communication Education, 28 (1979), 129-133.

[26]G.L. Piché, "Class and Culture in the Development of the High School English Curriculum, 1880-1900," Research in the Teaching of English, 11 (1977), 17-27.

[27]Mead, p. 20.

[28]G. Philipsen, "Speaking 'Like a Man' in Teamsterville: Cultural Patterns of Role Enactment in an Urban Neighborhood," Quarterly Journal of Speech, 61 (1975), 13-22.

[29]F. Williams and R.C. Naremore, "On the Functional Analysis of Social Class Differences in Modes of Speech," Speech Monographs, 36 (1969), 77-102.

[30]S.V. Philips, "Acquisition of Rules for Appropriate Speech Usage," Monograph Series on Language and Linguistics, No. 23 (Washington: Georgetown University Press, 1970), 77-96.

[31]For example, R.E. Bassett, N. Whittington, and A. Staton-Spicer, "The Basics in Speaking and Listening for High School Graduates: What Should Be Assessed?" Communication Education, 27 (1978), 293-303. For an inciteful attack on communication life skills approaches to English education see J.J. Rouse, "Knowledge, Power, and the Teaching of English," College English, 40 (1979), 473-491.

[32]Speaking order effects may pose a serious defect in this procedure. Data are currently being analysed to determine if this is so.

[33]R. Lloyd-Jones, "Primary Trait Scoring," in Evaluating Writing eds. C.R. Cooper and L. Odell (Urbana, Ill.: National Council of Teachers of English, 1977), 33-66.

Figure 1
Student Testing Materials

AGENDA

Custer County Board of Commissioners

The Custer County Board of Commissioners will hear citizens' comments about three proposals. Citizens who wish to speak to the County Commissioners may choose only one proposal. Your comments should be limited to three minutes. Take a few minutes now to think about what you want to say to the Commission. Prepare some notes or an outline to take with you when you speak.

1. Constructing a regional airport in Custer County:

   The proposed new airport will serve mostly business and industry. It will bring new jobs to our community. In order to build the new airport, the county will have to take out loans and raise taxes slightly. But the Federal government will pay half the cost. If enough flights use the airport, it will pay for itself in the future. The airport will create some noise and air pollution. The proposed site for the new airport is in an agricultural area two miles from the Custer County High School and five miles from downtown Custer.

2. Changing the present zoning regulations:

   Presently all areas of Custer County are zoned in one of four ways. "A" zones are for agricultural use only. "C" zones are for commerical use including stores and small businesses. "I" zones are the only permissible locations for industrial and manufacturing factories. "R" zones are residential, for private homes and apartment houses. This proposal eliminates all zoning regulations. With no zoning, people will be able to live closer to their work. More areas will be opened up for growth of housing, business, and industry. Homeowners and businesses will be

able to locate wherever they choose. Property values in some areas will rise. In other areas, quiet residential neighborhoods will be disturbed.

3. Establishing a County youth center:

The proposed youth center will serve residents of Custer County ages ten to eighteen. The center will include recreational facilities such as tennis and basketball courts, a swimming pool, and art studios. It will have rooms to be used by youth clubs free of charge. The youth center will also house a counseling service to help teen-agers with emotional and health problems. In order to pay for the new center, the County Commission will have to raise property taxes by 10%. For a home valued at $30,000 property taxes will increase by $25 per year.

Figure 2

PUBLIC HEARING FEEDBACK FORM

STUDENT:_____ DATE:_____

CLASS:_____ RATER:_____

SCORE:_____ PERFORMANCE STANDARD:_____

INTRODUCTION: (1) none; (2) just names proposal; (3) names proposal and attempts to capture interest; (4) names proposal and provides novel approach

PURPOSE: (1) no point of view; (2) vague point of view; (3) just states position on proposal; (4) states position with emphasis or situational qualifier

REASONS: (1) unsupported assertion; (2) unelaborated reasons given; (3) at least one reason supported; (4) at least on reason adapted to Commission's perspective

EVIDENCE (includes common knowledge): (1) no information or obviously incorrect information; (2) information of questionable validity or relevance; (3) clear and pertinent information; (4) information likely to be new or or interest to Commission

ORGANIZATION: (1) ideas wholly unrelated; (2) ideas implicitly or thematically related; (3) logical sequence or simple transitions; (4) proper emphasis and explicit connections between ideas

OBJECTIONS: (1) does not acknowledge reservations; (2) acknowledges but does not refute reservations; (3) refutes at least one reservation; (4) refutes most important reservation from Commission's point of view

CONCLUSION: (1) no conclusion or merely states that remarks are finished; (2) just thanks Commission; (3) restates position and offers thanks; (4) summarizes or concludes memorably and offer thanks

LANGUAGE STYLE: (1) slang or incomprehensible on several occasions; (2) very vague or distracting "fillers"; (3) fluent, appropriate formality; (4) vivid phrasing, highly comprehensible

GESTURES: (1) distracting mannerisms or posture; (2) no eye contact with Commissioners; (3) eye contact established, comfortable posture; (4) facial, body, or hand gestures used for emphasis or illustration

ORAL EXPRESSION: (1) monotone; (2) inappropriate or distracting inflection on several occasions; (3) natural variation in inflection; (4) tone of voice expresses conviction or emphasis

SPEECH RATE AND VOLUME: (1) inaudible; (2) rate too fast or too slow - distracting; (3) natural speech rate and volume; (4) variation in rate or volume used for added expressiveness